Info 206: Computing

Lecture 6

Networking and Internet Freedom

September 22, 2014

Random Walks Over the Web

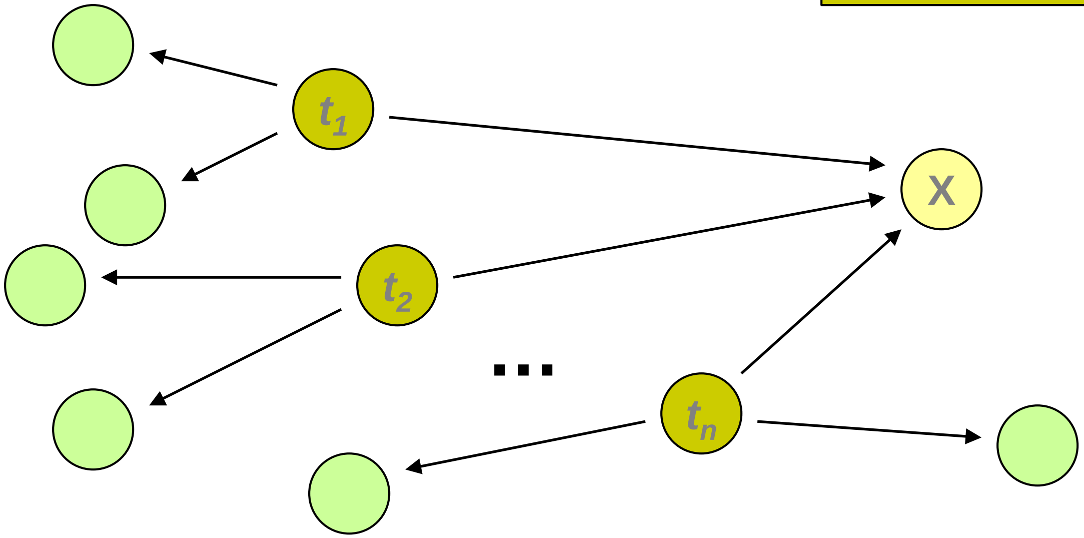
- Model:
    - User starts at a random Web page
    - User randomly clicks on links, surfing from page to page
- How much time is spent on each page?
- This is PageRank (named after Larry Page)

PageRank: Defined

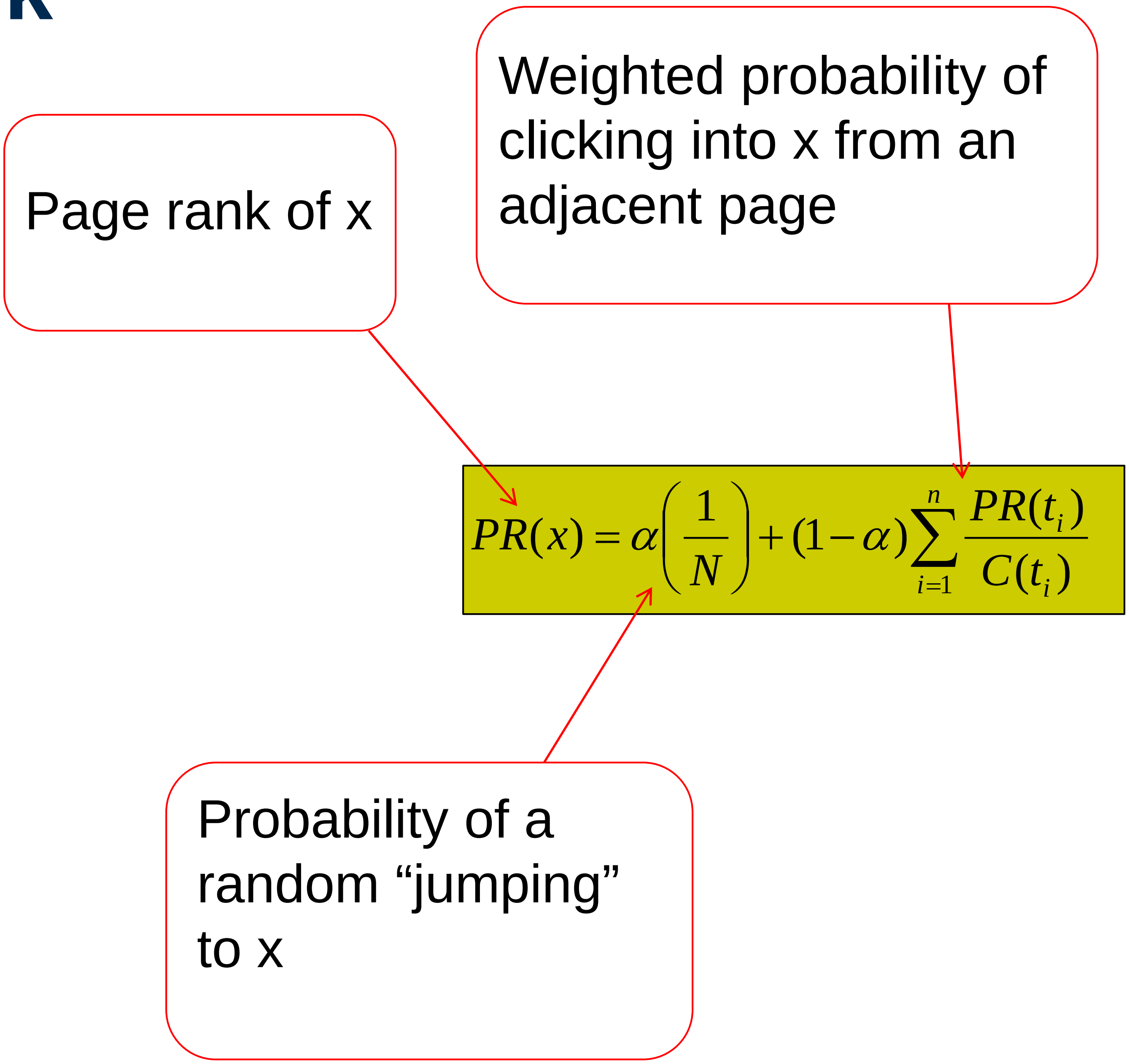
Given page *x* with in-bound links $t_1 \ldots t_n$, where

- $C(t)$ is the out-degree of t
- α is probability of random jump
- N is the total number of nodes in the graph

$$PR(x) = \alpha\left(\frac{1}{N}\right) + (1-\alpha)\sum_{i=1}^{n}\frac{PR(t_i)}{C(t_i)}$$



PageRank



Page rank of x

Weighted probability of clicking into x from an adjacent page

$$PR(x) = \alpha\left(\frac{1}{N}\right) + (1-\alpha)\sum_{i=1}^{n}\frac{PR(t_i)}{C(t_i)}$$

Probability of a random “jumping” to x

Computing PageRank

- Properties of PageRank
  - Can be computed iteratively
  - Effects at each iteration is local
- Sketch of algorithm:
  - Start with seed PR_i values
  - Each page distributes PR_i “credit” to all pages it links to
  - Each target page adds up “credit” from multiple in-bound links to compute PR_{i+1}
  - Iterate until values converge

Understanding Networking: OSI Model

	Layer Name	Description	Examples
➡	Application	User Level Processing	HTTP, FTP, Mail
	Presentation	Data Representation & Syntax	ISO Presentation
	Session	Sync Points and Dialogs	ISO Session
➡	Transport	Reliable End to End	TCP
➡	Network	Unreliable Thru Multi-Node Network	X.25 Pkt, IP
	Link	Reliable Across Physical Line	LAPB, HDLC
	Physical	Unreliable Wire, Telco Line	RS232, T1, 802.x

Understanding Networking: OSI Model

Layer Name	Description	Examples
Application	User Level Processing	HTTP, FTP, Mail
Presentation	Data Representation & Syntax	ISO Presentation
Session	Sync Points and Dialogs	ISO Session
Transport	Reliable End to End	TCP
Network	Unreliable Thru Multi-Node Network	X.25 Pkt, IP
Link	Reliable Across Physical Line	LAPB, HDLC
Physical	Unreliable Wire, Telco Line	RS232, T1, 802.x

IP/TCP

- Backbone protocol of Internet
- Internet Protocol
  - Moves packets from one location to another
- Transmission Control Protocol
  - Assures reliable reconstruction of data
  - Packets in order, no missing packets

IP version 4

- Addresses on Internet defined by four bytes
- ischool.berkeley.edu = 128.32.78.26
  - Handles many communications simultaneously
  - Needs “ports” to disambiguate
  - Port 22: SSH (secure shell)
  - Port 23: Telnet
  - Port 80: HTTP (web)
- Mapping from “domain names” to IP
  - Domain Name Service

TCP key issues

- Reliable communications
- Packets guaranteed to arrive in correct order

IP/TCP communication

- Defined by five values
  - Source IP address
  - Source port
  - Destination IP address
  - Destination port
  - Protocol (TCP is the most important one)

Other information

- Time to live
  - How long a packet can survive
  - Tracert
- Sequence numbers
  - Packets arrive in order
- Acknowledgement numbers
- Checksums
  - Make sure packet data uncorrupted

TCP handshake

A → B: SYN (synchronize)

B → A: SYN-ACK (synchronize & acknowledge)

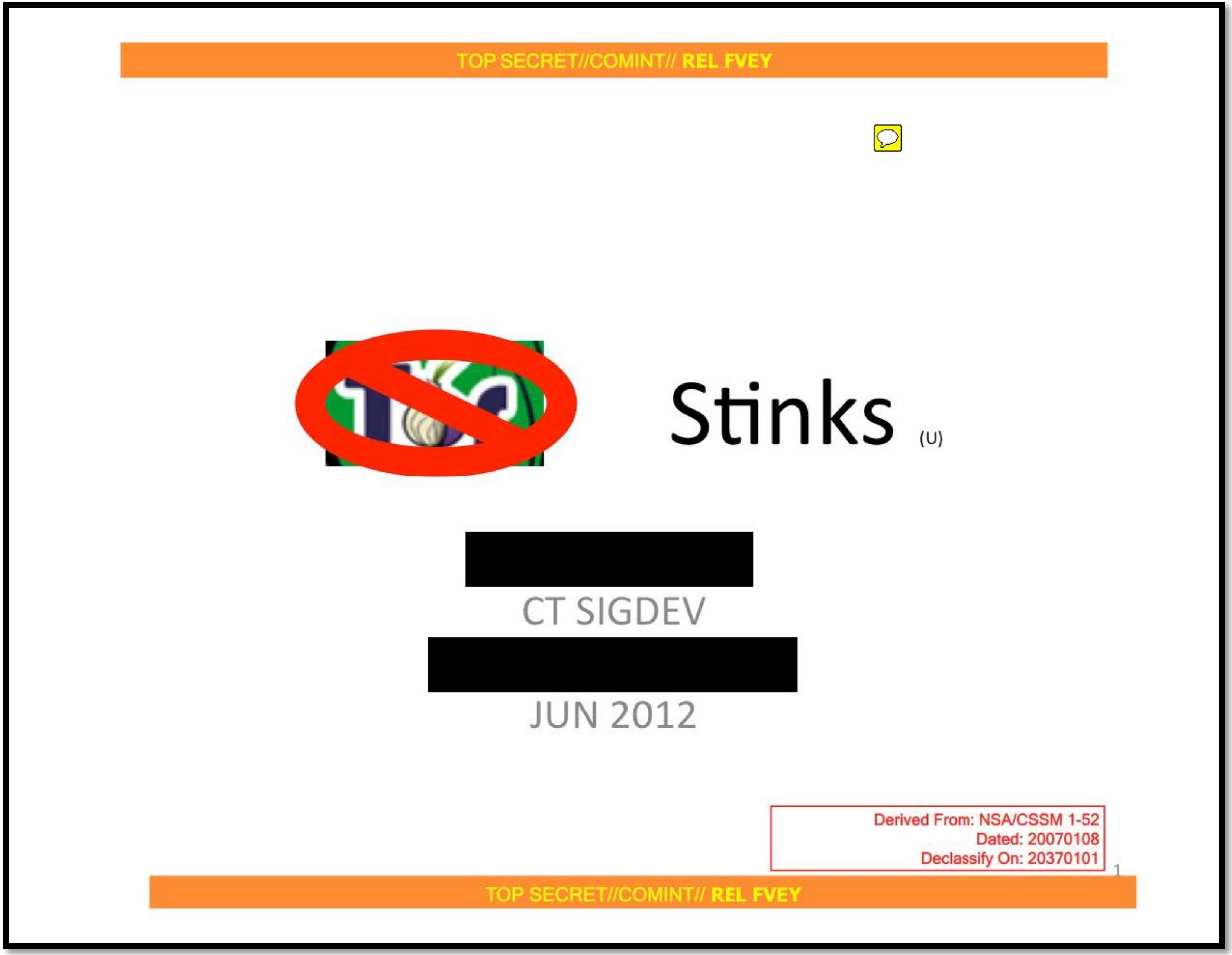
A → B: ACK (acknowledge)

Backbone protocol of the Internet

Internet Freedom

- Some countries censor access to Internet
  - Bahrain, Belarus, China, Cuba, Ethiopia, Iran, North Korea, Oman, Pakistan, Qatar, Saudi Arabia, Sudan, Syria, Turkmenistan, UAE, Uzbekistan, Vietnam, Yemen
- To address these concerns, users turn to *circumvention* programs (such as Tor, Freegate, Ultrasurf)

NSA: Tor stinks



TOP SECRET//COMINT// **REL FVEY**

Stinks (U)

[redacted]

CT SIGDEV

[redacted]

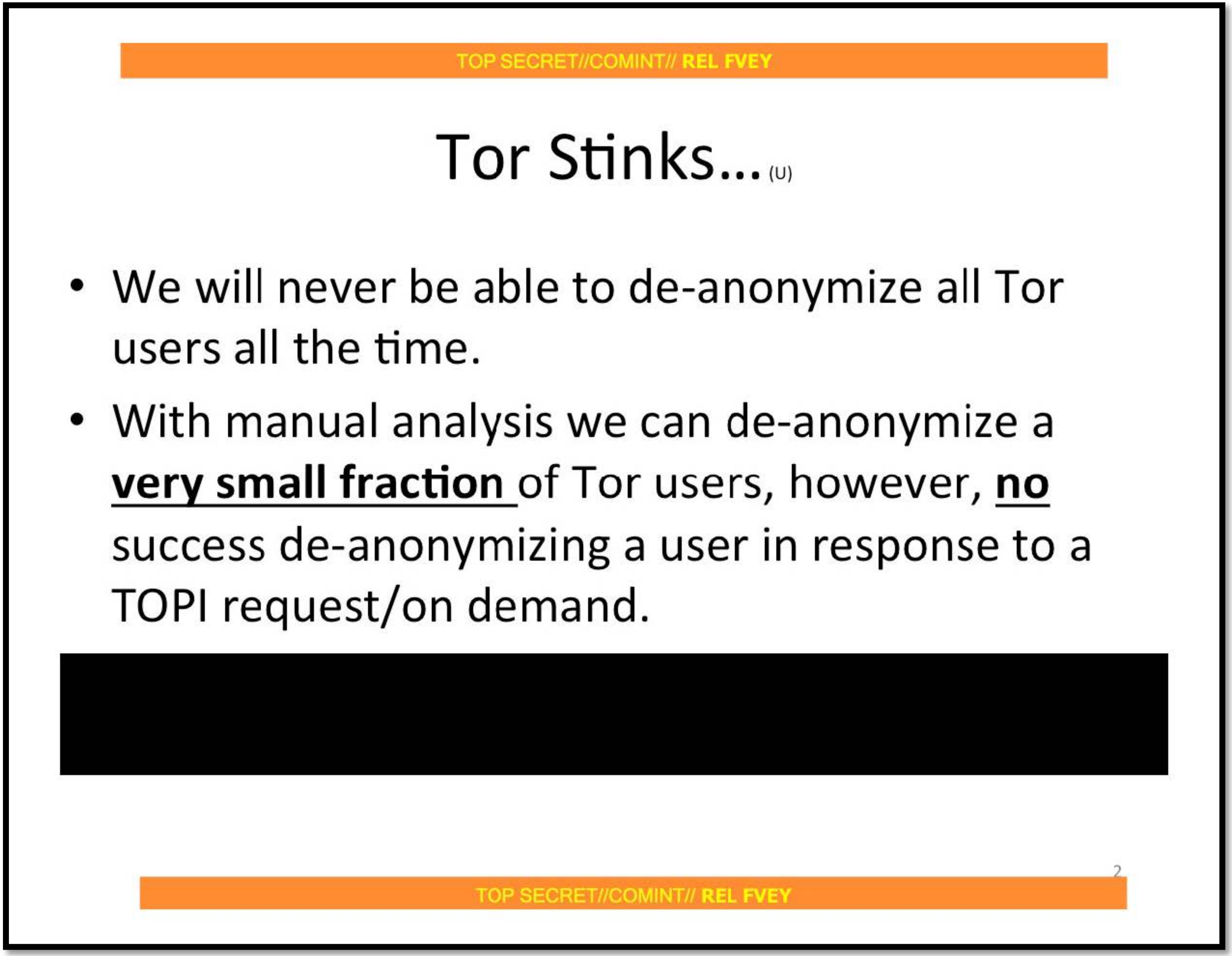
JUN 2012

Derived From: NSA/CSSM 1-52
Dated: 20070108
Declassify On: 20370101

1

TOP SECRET//COMINT// **REL FVEY**

NSA: Tor stinks



TOP SECRET//COMINT// **REL FVEY**

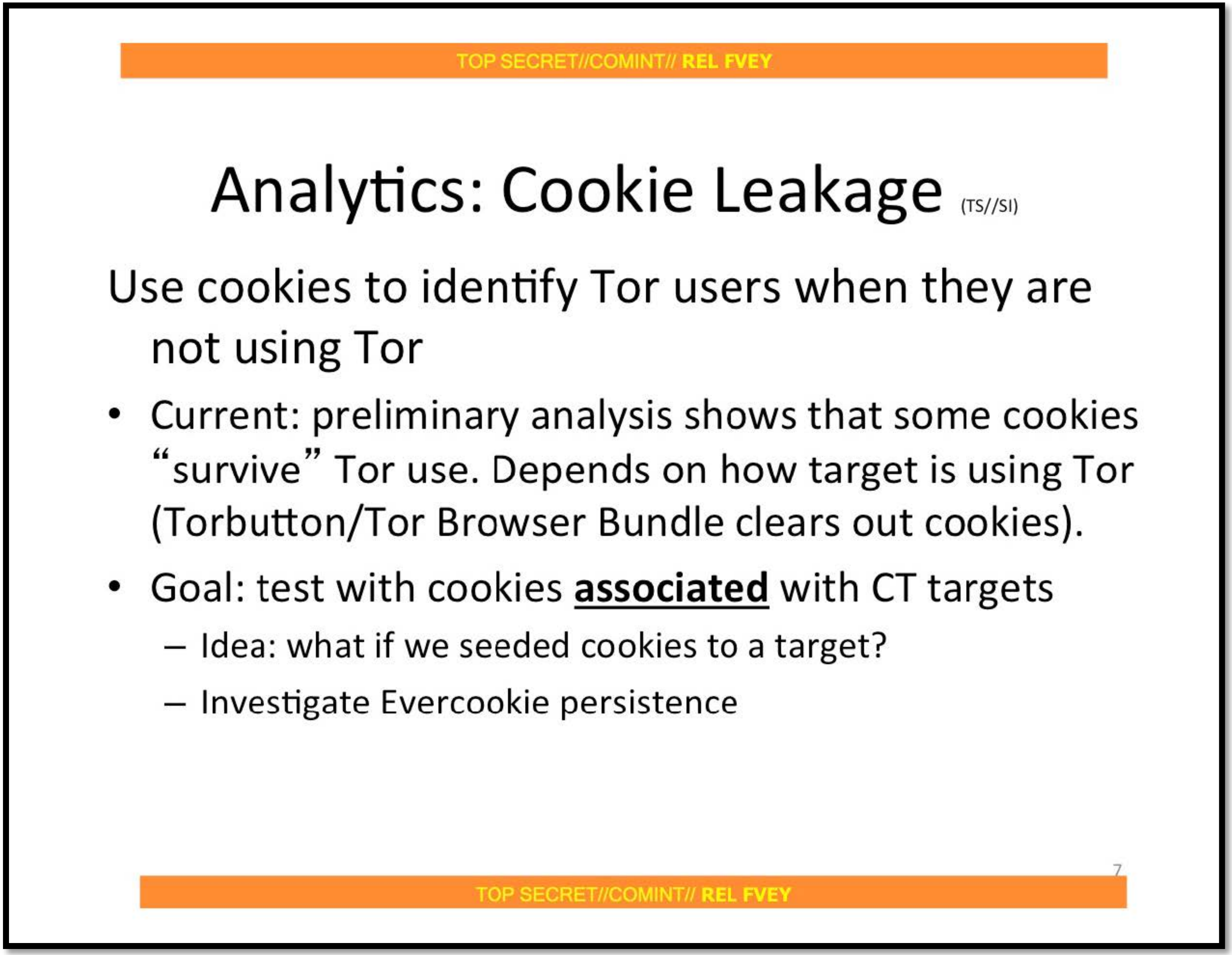
Tor Stinks… (U)

- We will never be able to de-anonymize all Tor users all the time.
- With manual analysis we can de-anonymize a **very small fraction** of Tor users, however, **no** success de-anonymizing a user in response to a TOPI request/on demand.

[redacted]

TOP SECRET//COMINT// **REL FVEY**

NSA: Tor stinks



TOP SECRET//COMINT// **REL FVEY**

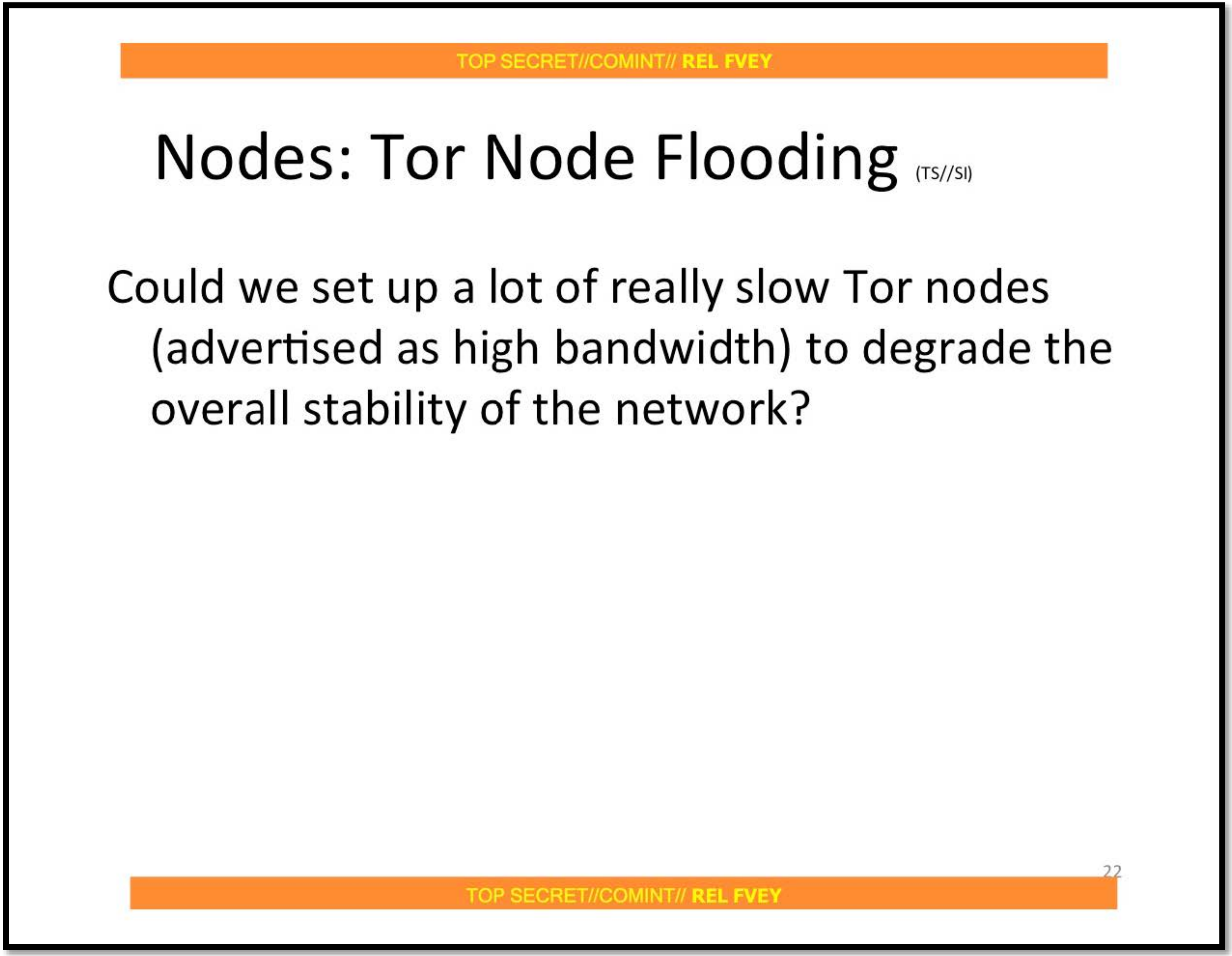
Analytics: Cookie Leakage (TS//SI)

Use cookies to identify Tor users when they are not using Tor

- Current: preliminary analysis shows that some cookies "survive" Tor use. Depends on how target is using Tor (Torbutton/Tor Browser Bundle clears out cookies).
- Goal: test with cookies **associated** with CT targets
  - Idea: what if we seeded cookies to a target?
  - Investigate Evercookie persistence

TOP SECRET//COMINT// **REL FVEY**

NSA: Tor stinks



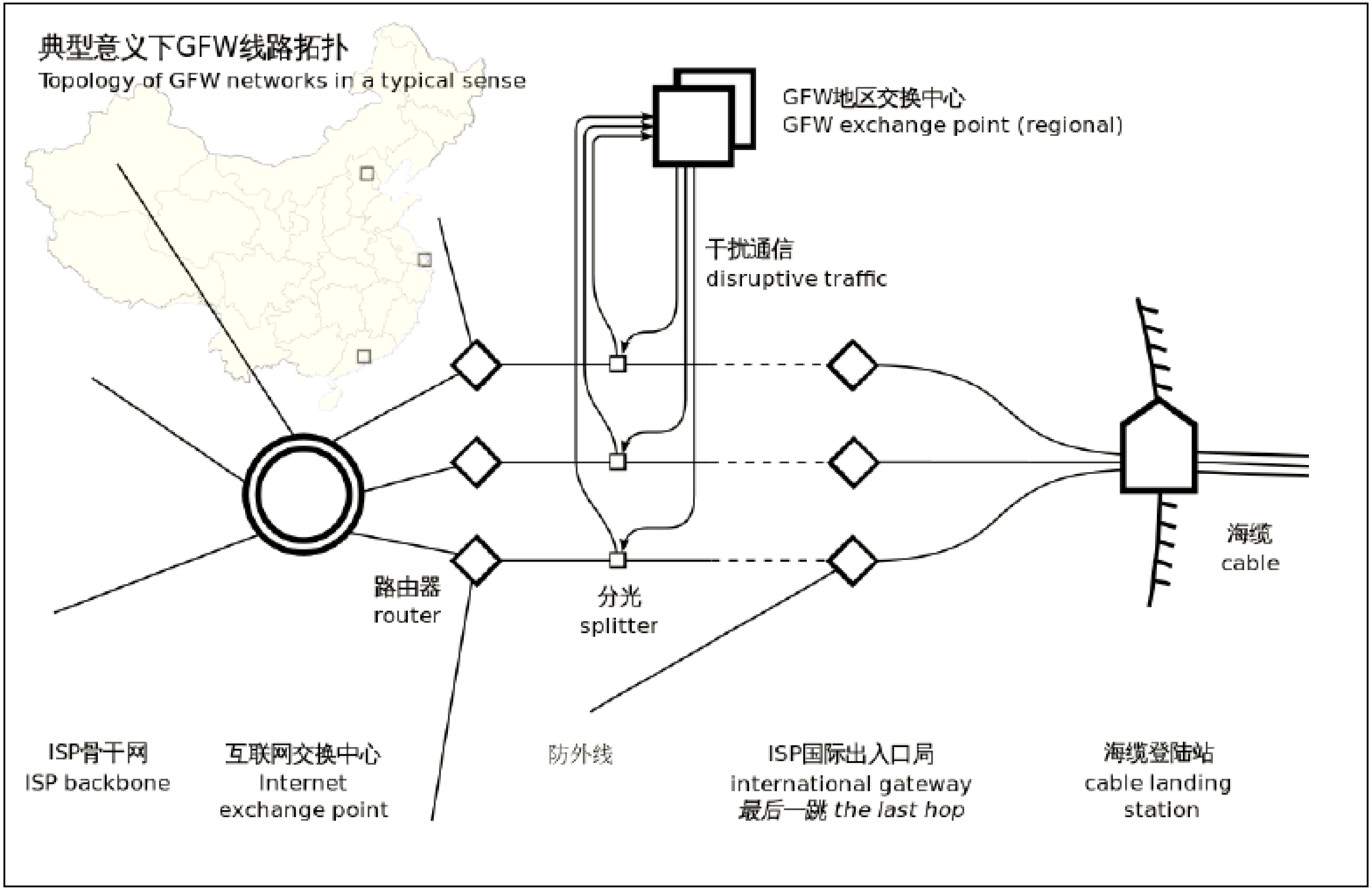
Nodes: Tor Node Flooding (TS//SI)

Could we set up a lot of really slow Tor nodes (advertised as high bandwidth) to degrade the overall stability of the network?

Basic censorship techniques

- Blacklisting IP addresses
    - ISP blocks a certain set of IP addresses
    - IP addresses may be circumvention tool
        - e.g., Tor nodes
    - IP addresses may be forbidden content
        - e.g., (dalailama.com)
- DNS poisoning
    - Mess up DNS lookup for certain domains
- Reset
    - Send a special TCP RST packet
    - Causes both sides to drop TCP connection

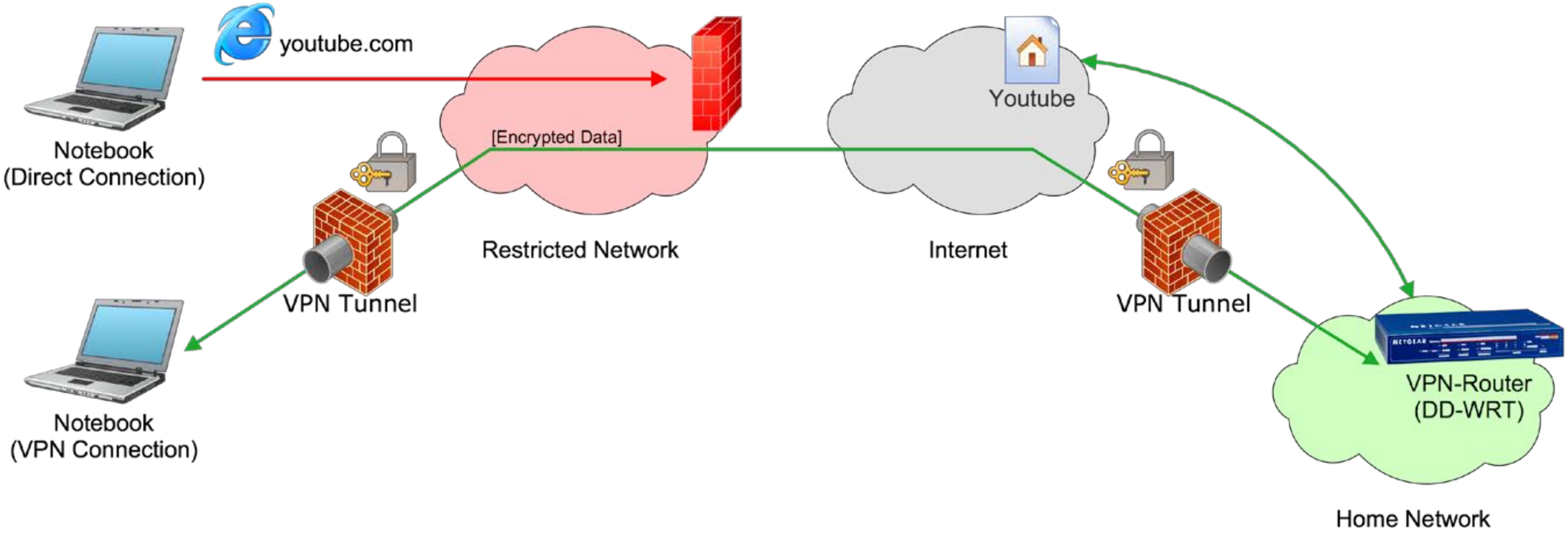
“Great Chinese Firewall”



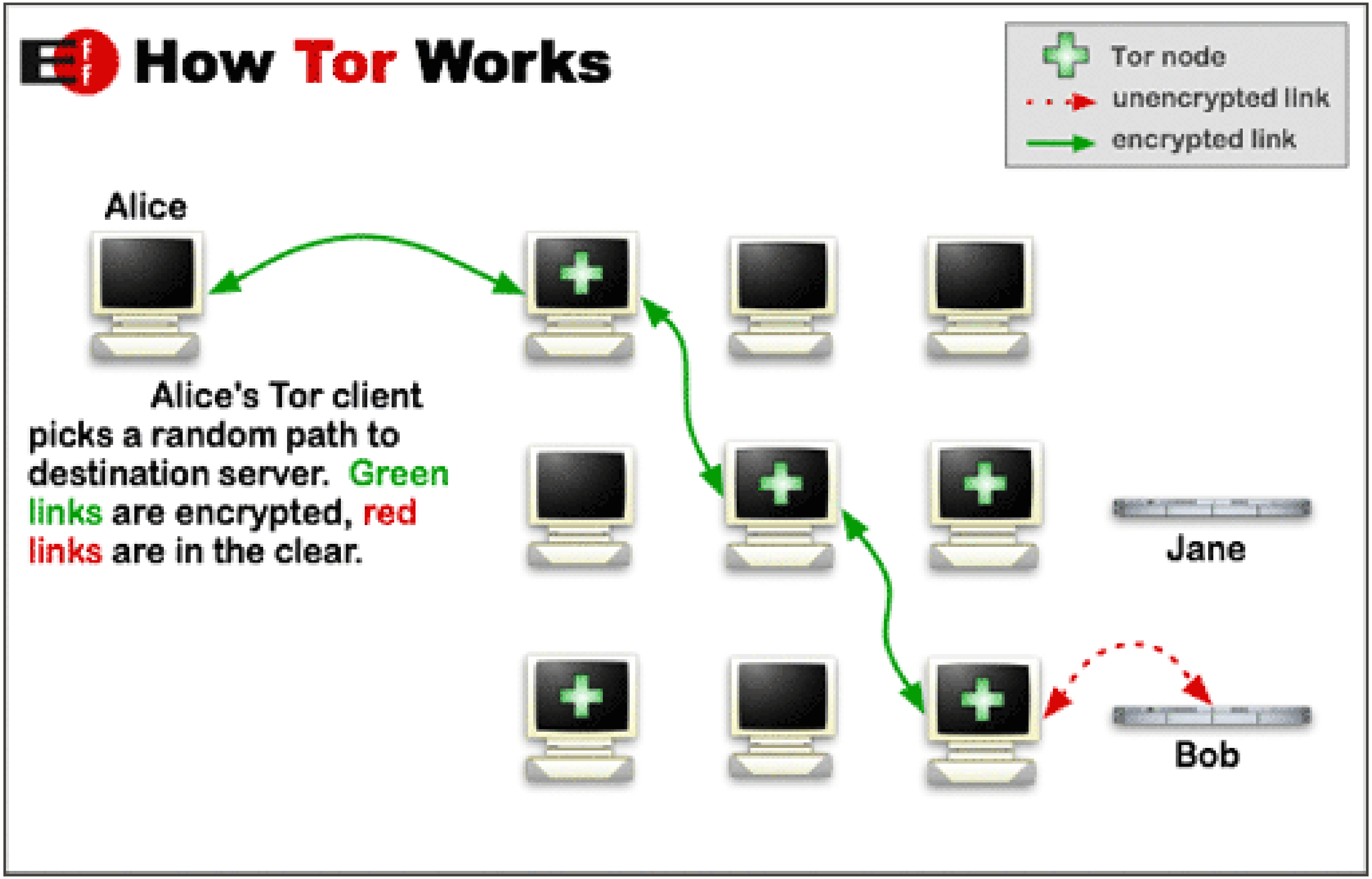
Circumvention needs

- Access
  - Just want to get to Youtube
- Privacy
  - Avoid surveillance
  - Key technology: encryption

One hop proxies/VPN



Multihop proxies (Tor)


How Tor Works
Tor node
unencrypted link
encrypted link
Alice
Alice's Tor client picks a random path to destination server. Green links are encrypted, red links are in the clear.
Jane
Bob

Tor usage

The anonymous Internet

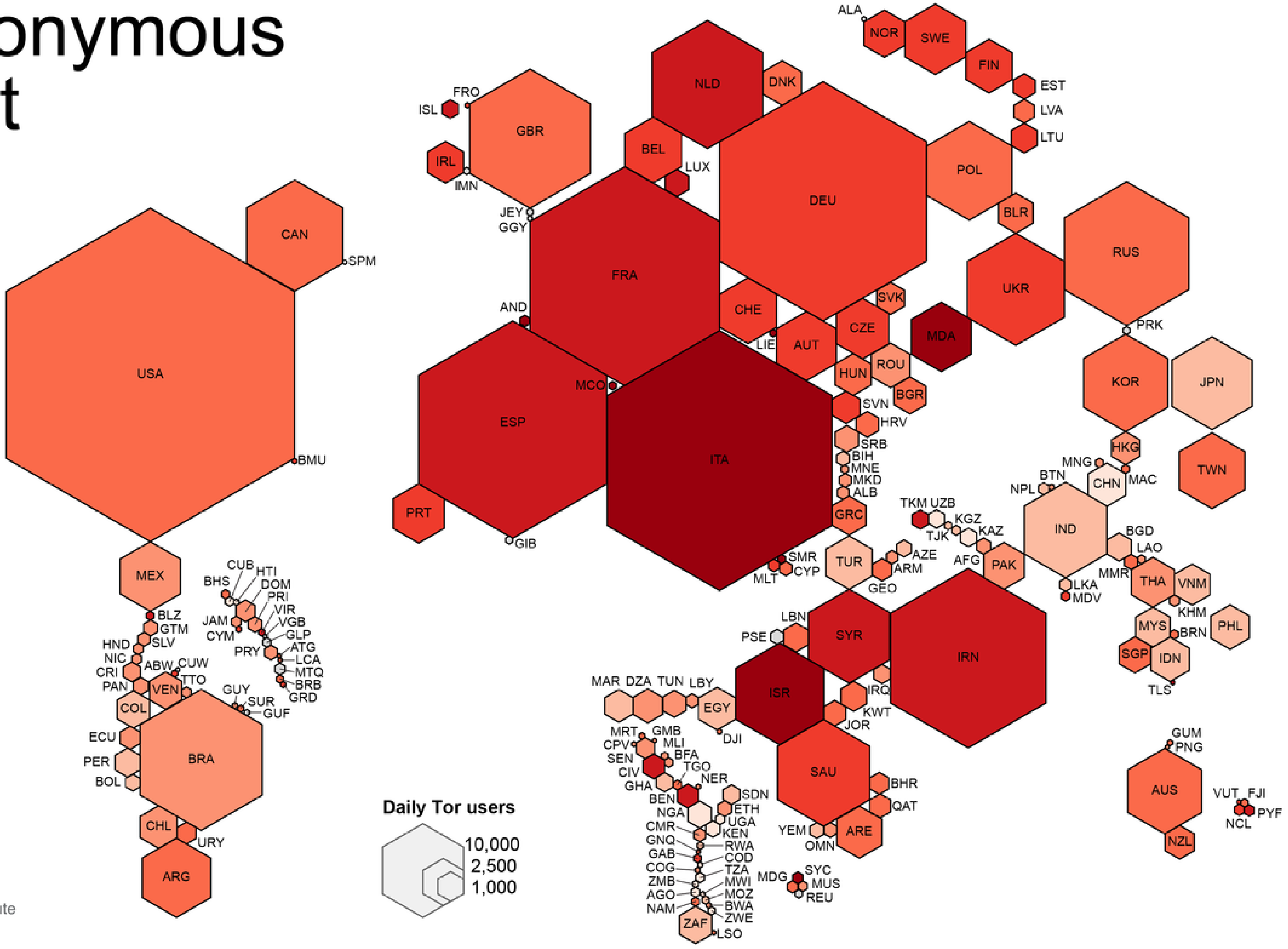
Daily Tor users per 100,000 Internet users

- 🟥 > 200
- 🟥 100 - 200
- 🟥 50 - 100
- 🟥 25 - 50
- 🟥 10 - 25
- 🟥 5 - 10
- □ < 5
- □ no information

Average number of Tor users per day calculated between August 2012 and July 2013

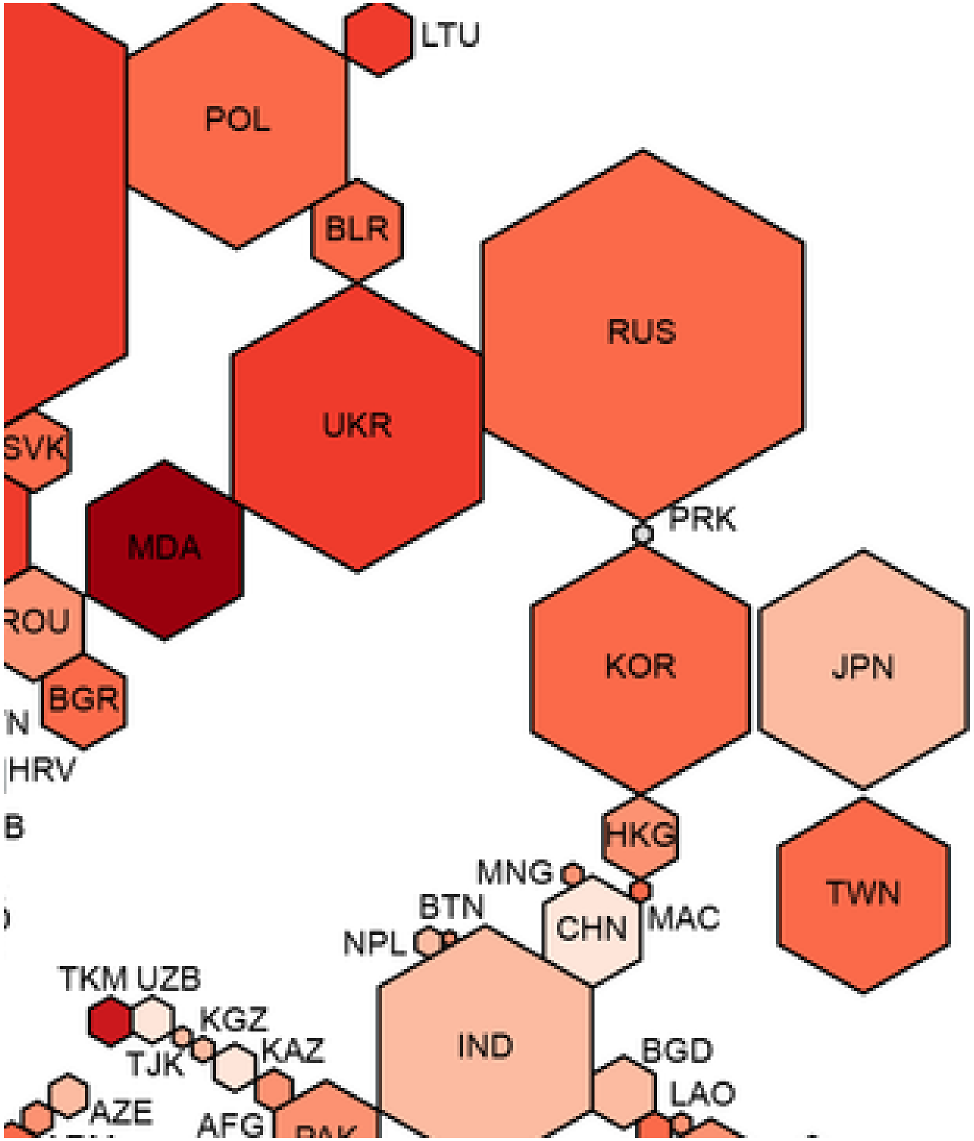
data sources:
Tor Metrics Portal
metrics.torproject.org
World Bank
data.worldbank.org

by Mark Graham (@geoplace) and Stefano De Sabbata (@maps4thought)
Internet Geographies at the Oxford Internet Institute
2014 • geography.oii.ox.ac.uk



Oxford Internet Institute
University of Oxford

Tor usage


LTU
POL
BLR
RUS
UKR
SVK
MDA
PRK
ROU
KOR
JPN
BGR
HRV
HKG
MNG
BTN
CHN
MAC
TWN
NPL
TKM
UZB
KGZ
TJK
KAZ
IND
BGD
AZE
LAO
AFG

Why isn’t Tor used more in China?

- Short answer: it doesn’t work
- Tor nodes (IP addresses) are discovered & blocked
- Tor protocol has “signature” allowing it to be blocked